IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Art Unit	:	1652
Examiner	:	Munjunath N. Rao
Applicant(s)	:	RL Henrikson, MJ Bienkowski
Serial Number	:	09/836,461
Filed	:	April 17, 2001
For	:	Heparanase II, A Novel Human Heparanase Paralog

Commissioner of Patents
P.O. Box 1450
Alexandria, VA 22313-1450

DECLARATION UNDER 37 C.F.R. 1.132

Sir:

I, Julie K. Lyons, the undersigned, a citizen of the United States and residing at 6797 Country View, Kalamazoo, MI 49009, declare and say as follows:

1. I was employed by the Pharmacia & Upjohn Company as a Legal Assistant on a date prior to April 18, 2000.

2 . It was my routine practice, during the ordinary course of business, during the relevant time period, to initial and date stamp Invention Disclosures which I received as part of my duties.

3. Applicant's attorney has shown me an Invention Disclosure entitled "Full-length cloning and characterization of a heparanase II, a novel human heparanase paralog" which is attached in redacted form as Exhibit A.

4. The Invention Disclosure is marked with my initials "JKL" in the upper right hand margin of the first page and is date stamped as was my routine practice during the relevant time period. ~~The actual date stamped on the Invention Disclosure,~~ the names of individuals not inventors of the claimed subject matter and certain other portions of the disclosure have been redacted.

5. The date stamp indicates that I received this document prior to April 18, 2000.

6. Exhibit A consists of:

A photocopy of the Invention Disclosure with portions of the Disclosure containing the date, identities of individuals and certain other portions of the disclosure redacted.

I hereby declare that all statements made herein are of my own knowledge, and I believe them to be true, and further that these statements are made with the knowledge that willful false statements so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Julie K Lyons

Julie K. Lyons

September 5, 2003

Date

Exhibit A

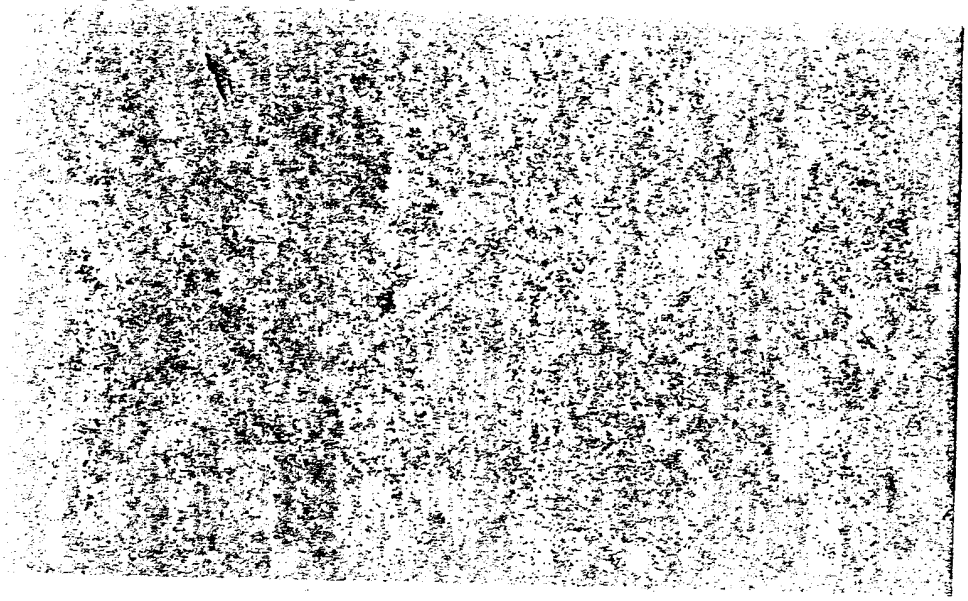
JKL → EFK

Full-length cloning and characterization of a heparanase II, a novel human heparanase paralog

6309

Robert L. Heinrikson, and Michael J. Bienkowski

1 SIGNATURE PAGE



Robert L. Heinrikson, Ph.D.
Protein Science
Pharmacia & Upjohn
Kalamazoo, MI USA

Michael J. Bienkowski, Ph.D.
Genomics Research
Pharmacia & Upjohn
Kalamazoo, MI USA

2 SUMMARY

The predicted amino acid sequence of human heparanase was used to interrogate genomic databases to identify paralogs of this important enzyme. A combination of cDNA and 5'RACE analysis defined a novel full-length human transcript referred to as heparanase II. Alignment of the heparanase I and heparanase II sequences revealed 41% shared identity at the amino acid sequence level. The heparanse II polypeptide contains a signal peptide and lacks predicted transmembrane segments, consistent with it being a secreted protein. Common motifs present in the predicted heparanase II sequence include canonical acceptor sites for N-linked glycosylation and phosphorylation by protein kinase C. Examination of the tissue distribution of expression of heparanase II transcripts by Northern analysis revealed a limited expression pattern with the highest levels in bladder,

prostate, and small intestine. Electronic transcript imaging in genomic databases confirmed the Northern data and revealed that heparanase II transcripts are most abundant in human tissues that are rich in vascular smooth muscle.

3 TABLE OF CONTENTS

4 INTRODUCTION

Regulated breakdown of the extracellular matrix (ECM) in many tissues is essential for embryonic development, morphogenesis, reproduction, and tissue resorption and remodeling. In pathological situations, degradation of the ECM is an obligatory step in both extravasation of inflammatory cells and metastisis of tumor cells. Degradation of the ECM requires the cooperative action of proteases (eg. matrix metaloproteases) and the endoglycosidase activity(s) that cleaves heparan sulfate chains referred to as heparanase (1). Heparan sulfate proteoglycans are important components of the ECM, serving both structural function and as a reservoir for multiple growth factors like bFGF (2,3). Degradation of heparan sulfate by heparanase activity both compromises the integrity of the ECM barrier and liberates heparan sulfate bound growth factors. Heparanase mediated turnover of the ECM represents an essential step in cell migration processes including inflammatory cell extravasation (4) and tumor cell metastisis (5) and the release of growth factors that are important mediators of wound healing and angiogenesis (6).

Given the central role of heparanase in both normal and pathophysiological processes, molecular definition of the heparanase polypeptide has recently been achieved (7-11). Heparanase activity was used to guide the purification of the enzyme from either human platelets or transformed human cell lines and peptide sequences derived from the purified polypeptide were employed to clone a full-length cDNA encoding human heparanase (7-11). Using this experimental paradigm, five different groups all identified the same heparanase polypeptide sequence and this polypeptide did not share significant homology with any known protein. Functional characterization of this polypeptide revealed that it required proteolytic processing for activity and that ectopic expression of the cDNA in mammalian cells significantly increased their metastatic potential (7). Despite the presence of heparanase sequence tags in the public domain expressed sequence tag (EST) databases, no paralogs of heparanase were identified in the public data (8).

The predicted amino acid sequence of human heparanase was used to interrogate the proprietary Incyte genomic databases for heparanase paralogs. In addition to ESTs with exact matches to the heparanase sequence, multiple ESTs with significant shared identity to heparanase were also identified. Full-length cDNA cloning of the transcript encoding these ESTs defined a related human enzyme referred to as heparanase II. Multiple polypeptide isoforms of heparanase II, presumably formed by alternative splicing of a common human gene, were identified. Expression of the heparanase II gene was highest in human tissues enriched in smooth muscle.

5 OBJECTIVES

To discover heparanase I paralogs as potential drug targets.

6 MATERIALS AND METHODS

6.1 Computer-assisted analysis of EST databases, cDNA , and predicted polypeptide sequences

Genomic database mining of Incyte [LifeSeq, LifeSeq FL, LifeSeq Assembled, LifeSeq Gold, and LifeSeq Atlas], GenBank, and the Institute for Genomic Research Total Human Consensus databases was performed using the BLAST search tool. Contig assemblies and Clustal W multiple sequence alignments were performed using the bioinformatics tools provided with the Incyte LifeSeq database interface. Protein motifs were identified using either the ProSite dictionary [motifs in GCG Version 9.0] or the Pfam database [P&U Sweden]. Analysis of the polypeptide sequences for the presence of signal sequences [SignalP, (12)], transmembrane segments [TMHMM, (13)], and canonical acceptor sites for O-linked glycosylation [NetOGlyc, (14)] was performed using the algorithms provided by the Center for Biological Sequence Analysis [CBS] server at the Technical University of Denmark [URL http://www.cbs.dtu.dk/services/].

6.2 Full-length cDNA cloning of Human Heparanase II

Routine queries of LifeSeq and LifeSeq-Assembled databases using the full-length heparanase I sequence initially identified a series of non-overlapping sequence tags derived from distinct cDNA libraries (1654352, 3207353, and 3704980). These cDNA clones were obtained from Incyte and plasmid DNA prepared by alkaline lysis and banding in CsCl. Each clone was completely sequenced by primer walking on both strands using automated cycle sequencing with fluorescent terminator dyes. The sequences of clones 1654352, 3207353, and 3704980 were then used to query the Incyte databases, resulting in the identification of clones 3529440 and 3385824, which contained additional sequence compared to the original cDNAs.

Additional 5' DNA sequence was established by 5' RACE analysis using a Marathon-ready cDNA template obtained from Clontech (Palo Alto, CA). An antisense primer specific for the shared 5' region of cDNAs 3207353 or 3385824 [GGCAACATCACTTCGAACAATGTC] was paired with the universal AP-1 primer in the PCR on a Marathon-ready cDNA templates prepared from either human prostate, human small intestine, human bladder, or human heart RNA (Clontech, Palo Alto, CA). The following thermocycle parameters were used;

1 min @ 94°C
30 sec @ 94°C, 4 min @ 72°C for 5 cycles

30 sec @ 94°C, 4 min @ 70°C for 5 cycles
30 sec @ 94°C, 4 min @ 68°C for 25 cycles
10 min extension @ 72 °C

Specific amplification products were not detected by agarose gel analysis of the primary 5' RACE products so a nested amplification was performed. The primary amplification products (5µl) were diluted with 245 µl water and 5 µl of the resulting mixtures taken for nested amplification. Primer AP-2 (Clonetech, Palo Alto, CA) was paired with the nested primer specific for the 5' end of clone 3207353 [CGAGCCAGCCATCATGAATGATG]/human prostate and human small intestine templates or specific for the 5' end of clone 3385824 [GAGAGGAAAGGTTCCCAGGACAG]/human bladder and human heart templates and PCR amplification performed exactly as described above. A single major amplification product was obtained in each case and the products were cloned into the *Sma*I site of pUC18 using the SureClone kit (Amersham Pharmacia Biotech, Arlington Heights, IL). Two isolates from each RNA source were completely sequenced on both strands by cycle sequencing.

6.3 Tissue distribution of expression of heparanase II transcripts

The tissue distribution of expression of human heparanase II was determined using multiple tissue Northern blots obtained from Clontech (Palo Alto, CA). Incyte clone 3704980 in the vector pINCY was digested to completion with *EcoRI*/*Not*I and the cDNA insert purified by preparative agarose gel electrophoresis. This fragment was radiolabeled to a specific activity > 1 X 10^9 dpm/µg by random priming in the presence of [α-^{32}P-dATP] (>3000 Ci/mmol, Amersham, Arlington Heights, IL) and Klenow fragment of DNA polymerase I (Amersham Pharmacia Biotech, Piscataway NJ). Nylon filters containing denatured, size fractionated poly A^+ RNAs isolated from different human tissues were hybridized with 2 X 10^6 dpm/ml probe in ExpressHyb buffer (Clontech, Palo Alto, CA) for 1 hour at 68 °C and washed as recommended by the manufacturer. Hybridization signals were visualized by autoradiography using BioMax XR film (Kodak, Rochester, NY) with intensifying screens at -80 °C.

7 RESULTS

Identification and full-length cloning of a heparanase paralog—Molecular definition of human platelet heparanase was recently achieved using a combination of protein sequencing and mining of expressed sequence tag databases (7-11). The predicted amino acid sequence of human heparanase was used to interogate the Incyte databases using the FASTA search tool. In addition to the ESTs displaying an exact match to the heparanase sequence, three additional ESTs were detected. Each of these ESTs showed approximately 40% shared identity with the heparanase amino acid sequence, consistent

with a paralog relationship. These three EST sequences could not be assembled into a contig, indicating that either they are derived from non-overlapping regions of a single gene or they are derived from as many as three separate human genes. To resolve this issue, Incyte clones 1654352 (prostate tumor library), 3207353 (corpus cavernosum), and 3704980 (corpus cavernosum) were obtained and completely sequenced on both strands to provide 100% accurate sequence. Subsequent queries of the Incyte databases with these cDNA sequences and the BLAST search tool identified several additional EST matches. Incyte clones 3529440 (normal bladder) and 3385824 (normal esophagus) were also obtained and completely sequenced. Alignment of the sequences of clones 1654352 (960 bp), 3385824 (2350 bp), 3529440 (3360 bp), and 3704980 (1384 bp) using the Clustal W algorithm is shown in Figure 1. The sequence of clone 3207353 was not included in this alignment because the 3' sequence diverged from all of the other cDNAs (see below). Clone 3385824 contained the 5'-most sequence and the sequences of clones 1654352 and 3704980 were co-linear with the sequence of 3385824. The sequence of clone 3529440 was also co-linear with the other sequences except that it contained an extension of > 1000 nucleotides in the 3' non-coding region. This is likely to be a result of alternate polyadenylation of the heparanase II transcript. Despite the large differences in the sequence lengths, all four cDNAs contained a region of greater than 900 nucleotides of shared identity. This confirmed that all four cDNAs were derived from transcription of a single human gene referred to as heparanase II. A portion of the sequence of clone 3207353 was co-linear with the other sequences (858 nucleotides) but it diverged on both the 5' and 3' ends. On the 5' end, the sequence was identical to 3385824 except for the 5'-most 162 nucleotides, which did not match the 110 nucleotides on the 5' end of clone 3385824. On the 3' end of clone 3207353, the sequence diverged from the other 5 clones 456 nucleotides upstream of the predicted translation termination codon. The divergent sequence contains an in-frame stop codon that deletes the C-terminal 152 amino acid residues and replaces them with the 4 amino acid residues shown in Figure 2. Also, the 3' region of clone 3207353 contains an *Alu* repeat. Since *Alu* repeats are almost exclusively found in non-coding regions and the sequence diverges from five other cDNAs in this region, we conclude that the 3' end of clone 3207353 is likely derived from a partially spliced transcript or a chimeric cDNA. A Clustal W alignment of the predicted amino acid sequences of all five cDNA clones is shown in Figure 2. Clones 3207353 and 3385824 contain the most up-stream sequence but diverge in the N-terminal region. The remaining clones have complete shared identity with clone 3385824 through the translation termination codon except for a polymorphism near the C-terminus (clones 3385824 and 3704980 contain Phe while clones 1654352 and 3529440 contain Tyr). Based on this analysis, none of the cDNAs were full-length as evidenced by the lack of an in-frame translation initiation codon.

The remainder of the coding sequence was determined by 5' RACE analysis. To confirm the alternative 5'-exon usage predicted from the cDNA analysis, a series of cDNA templates were amplified in the PCR on cDNA templates using oligonucleotide primer pairs specific for either 3207353 or 3385824. An amplicon of the expected size for clone 3207353 was observed in templates derived from human prostate, human small intestine, human bladder and human heart. In contrast, an amplicon specific for clone 3385824 was

only observed in templates derived from human bladder and human heart (data not shown). These results confirm that these 5'-alternative splice variants were not cloning artifacts real and that they show tissue specific expression. For 5' RACE analysis, marathon-ready cDNA from these four human tissues was amplified in the PCR using a universal AP-1 sense primer paired with a gene-specific antisense primer that was designed from the common region of clones 3207353 and 3385824. Analysis of the products obtained from both templates did not reveal specific product(s). A round of nested amplification, using the universal AP-2 sense primer paired with an antisense primer specific for clone 3207353 (small intestine and prostate) or 3385824 (bladder and heart), was then performed. The nested amplification gave an excellent yield of specific products from either nested primer pair and this material was subcloned and sequenced. Combining the sequence derived from 5' RACE analysis of clone 3207353 with the sequences assembled from the cDNA clones yielded a composite full-length sequence for heparanase II (Figure 3). The composite cDNA contained a 1602 bp open-reading frame that encoded a novel polypeptide containing 534 amino acid residues. Alternatively, the 5' RACE analysis of the 3385824 transcripts did not yield any additional sequence information beyond the original clone.

Computer-aided analysis of the predicted heparanase II amino acid sequence—The predicted amino acid sequence of human heparanase II was analyzed for various protein motifs using both the ProSite dictionary and the Pfam database as well as using prediction methods available on the Center for Biotechnology Sequence Analysis (CBS) server in the Biotechnology Department at the University of Denmark. The ProSite motifs analysis identified canonical acceptor sites for Asn-linked glycosylation [alignment positions 217 and 334] and consensus acceptor sites for phosphorylation by protein kinase C [alignment positions 66, 107, 116, 163, 218, 323, 330, and 350]. Also, a potential sites for C-terminal amidation [G-R/K-R/K] were localized to alignment positions 116 and 315. The heparanase II amino acid sequence was analyzed for the presence of a signal sequence using the SignalP neural net-based prediction method available on the CBS server. Using neural nets trained on eukaryotic signal sequences, the first 41 NH_2-terminal amino acid residues are predicted to be a signal peptide based on all four parameters and the most likely site of cleavage is between positions 41 and 42 [SQA↓GD]. No predicted transmembrane domains were detected in the human heparanase II sequence. The presence of a signal sequence and the lack of predicted transmembrane segments are consistent with heparanase II being a secreted protein. The positions of these various functional motifs in the heparanase II amino acid sequence are summarized in Figure 4.

The sequence of human heparanase II was then aligned with the predicted sequence of heparanase I using the Clustal W algorithm and the results are shown in Figure 5. Heparanase I and II display 43% shared identity at the amino acid sequence level with 213 identical residues.

Tissue distribution of expression of heparanase II—The tissue distribution of expression of human heparanase II was established using a combination of Northern blot analysis

and electronic querying of the Incyte databases. For Northern analysis, heparanase II transcripts were visualized using a cDNA probe derived from Incyte clone 3704980 and the results are shown in Figure 6. A single 4.4 kb transcript was detected at the highest level in bladder and lower amounts were also present in prostate, stomach, small intestine, uterus and brain. No signal was detected in skeletal muscle, colon, heart, thymus, spleen, kidney, liver, placenta, lung, or peripheral blood leukocytes under these conditions (data not shown). A BlastN search of the LifeSeq-Gold database using the full-length cDNA sequence for human heparanase II revealed a total of 14 exact matches in four distinct gene templates. The library source of these EST sequences is summarized in Table 1. Template 273691.1 contains two clones, both derived from tumors (prostate, 1654352 and breast, 3775436). A survey of the clones that populate the other three gene templates reveals that the majority of the clones were sequenced from tissues rich in vascular smooth muscle (*eg.* corpus cavernosum, esophagus, bladder, femoral artery, and uterine cervix). Taken together, the results of both the Northern analysis and database searches confirm that heparanase II expression is peculiar to tissues that are rich in vascular smooth muscle.

8 DISCUSSION

The LifeSeq-Gold database contains four distinct assemblies with significant shared identity to the human heparanase I sequence. Systematic analysis of cDNA clones contained within these gene bins confirmed that they are all derived from transcription of a single human gene, referred to as heparanase II. Both heparanase I and heparanase II share a similar domain organization including a relatively long signal peptide followed by a catalytic domain that lacks predicted transmembrane segments. This organization is consistent with both heparanases I and II being secreted proteins. Motifs shared by both polypeptides include canonical acceptor sites for N-linked glycosylation, phosphorylation by protein kinase C, and for C-terminal amidation. The consensus sites for tyrosine phosphorylation and PKA in heparanase I are not present in heparanase II.

The predicted amino acid sequence of heparanase II does not show significant identity to any protein in the November 1999 release of SwissProt, except heparanase I. The availability of two related polypeptide sequences with little homology to other known proteins allows predictions to be made reagrding structure-function. Assuming that the heparanase I and II genes arose by duplication and subsequent divergence of a single ancestoral gene, regions of the polypeptide sequence important for function are likely to be conserved. For example, heparanase I is initially synthesized as pro-heparanase I that is proteolytically processed into a two chain heterodimer (11). Alignment of the human heparanase I and human heparanase II amino acid sequences (Figure 5) revealed that only one of the two processing sites is conserved. The processing sites in heparanase I involve the excision of a 44 or 45 amino acid region near the N-terminus by sequential proteolytic cleavage at the sequence PKK↓EST or PKKE↓ST and HYQ↓KKF to generate the N-

terminus and C-terminus of the excised peptide, respectively. By alignment, the predicted processing sites in heparanase II would be NLR↓NPA and DKQ↓KGC, indicating conservative substitutions in the N-terminal P1/P1' positions and identical P1/P1' residues at the C-terminal processing site. Whether hepananase II is processed to a two chain heterodimer is unknown at present.

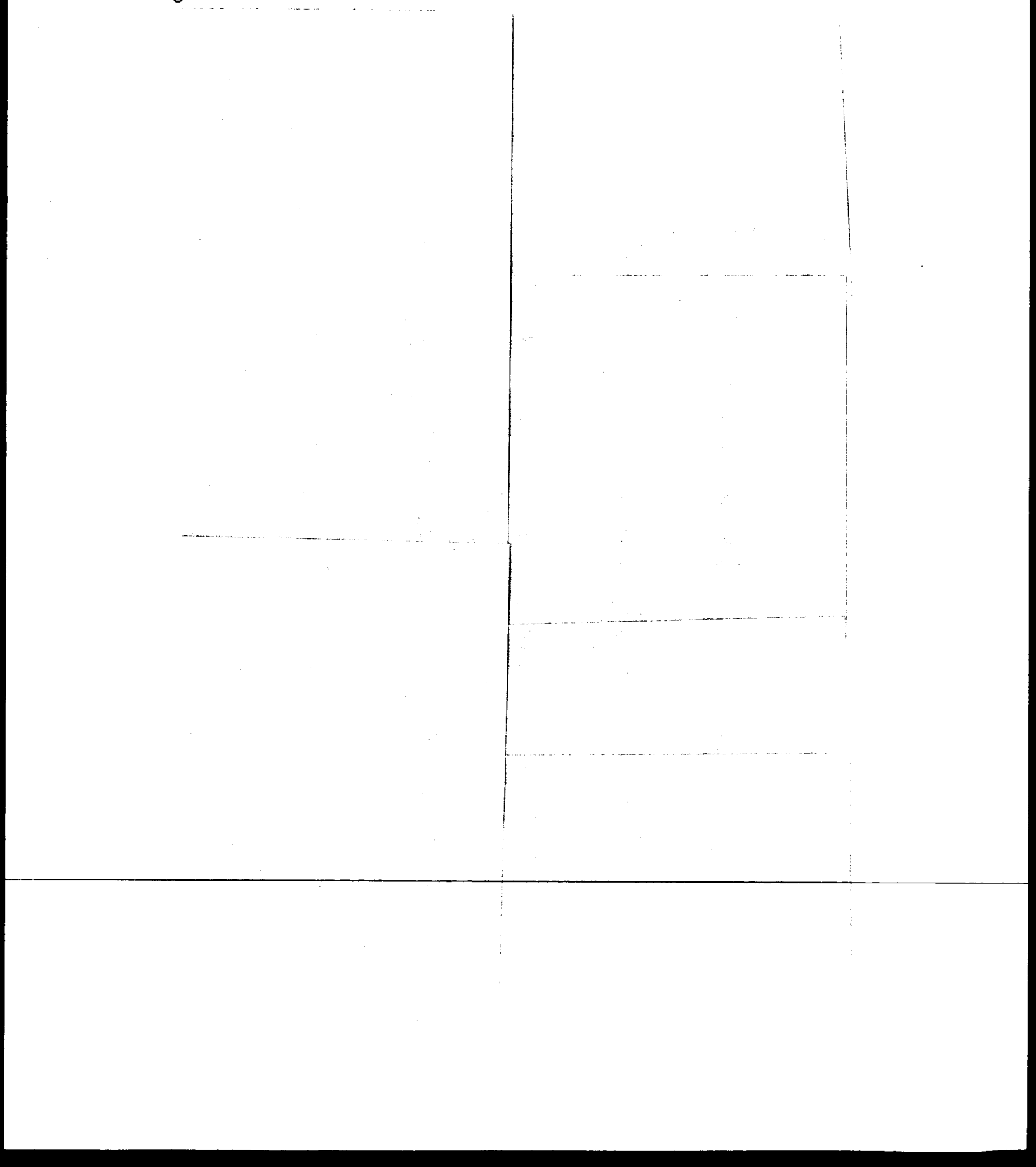
Examination of the enzymatic activity of native platelet heparanase I has revealed that the enzyme is an endo-β-glucuronidase (15). The enzyme preferentially cleaves heparan sulfate between D-glucuronic acid and N-acetylglucosamine residues in which the uronic acid on the reducing side of the N-acetylglucosamine is O-sulfated (15). Glycosidases function by two general mechanisms resulting in either retention or inversion of configuration at the hydrolysis site (16, 17). In both cases, two acidic amino acids, usually glutamic acids, are directly involved in catalysis. The acidic side chain of one amino acid serves as the nucleophile while the other acts as a general acid/general base in the reaction mechanism. Structure-function studies of lysosomal human exo-β-glucuronidase involved in the degradation of glycosaminoglycans implicates a pair of glutamic acid residues (Glu^{451} and Glu^{540}) in the catalytic mechanism (18, 19). Alternatively, the catalytic pair in lysozyme involves Glu^{35} and Asp^{52} (20). Taken together, these results suggests that a pair of conserved amino acid residues with acidic side chains in heparanase I and II may participate in the endo-β-glucuronidase activity of both enzymes. Inspection of the Clustal W alignment of the heparanase I and II amino acid sequences revealed 15 aspartic acid residues that are conserved between the two sequences but no glutamic acid residues. Six of these aspartic acid residues are nested in clusters of sequence identity that involve >75% identity over >15 amino acid residues. One or more of these regions are likely to contribute the residues involved in heparanase catalysis.

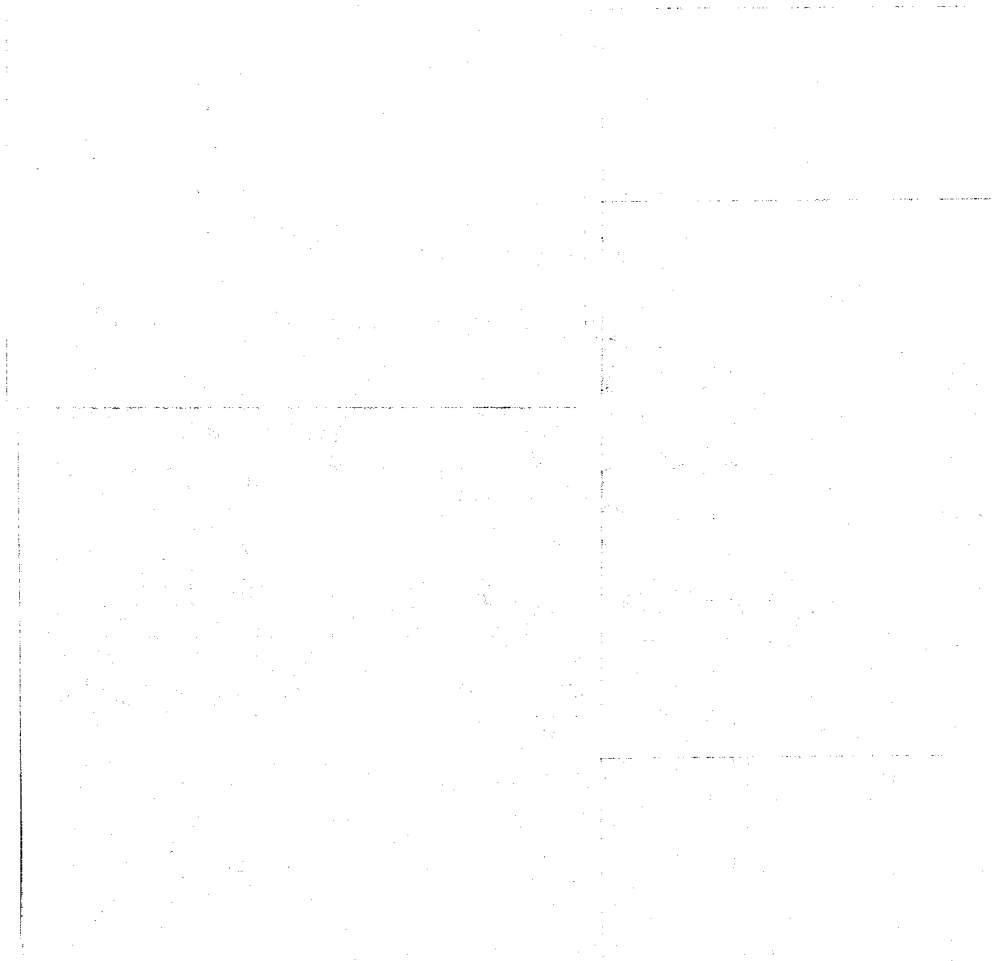
In addition to distant relationship regarding sequence, another remarkable distinction between human heparanase I and human heparanase II is related to their tissue distribution of expression. Common sources of heparanase activity include human platelets, placenta, and tumor cell lines and the enzyme from both platelets and tumor cell lines are biochemically indistinguishable (15, 21). Indeed, Northern blot analysis of the human tissue distribution of expression of heparanase I revealed high expression levels in placenta and peripheral blood leukocytes and somewhat reduced levels in spleen, lymph node, bone marrow and fetal liver (8,10). We could not detect the expression of heparanase II in placenta, peripheral blood leukocytes (data not shown) but rather observed the highest level of expression in tissues rich in vascular smooth muscle (Figure 6). A survey of the expression pattern of heparanase enzyme activity has identified vascular smooth muscle cells as a source of activity (22). These results indicate that heparanase I and II have a non-overlapping expression pattern in human tissues and each may serve tissue-specific functional roles. Perhaps the substrate specificity for heparan sulfate hydrolysis is distinct between these two isozymes and the work reported here enables the preparation of recombinant heparanase II for further characterization.

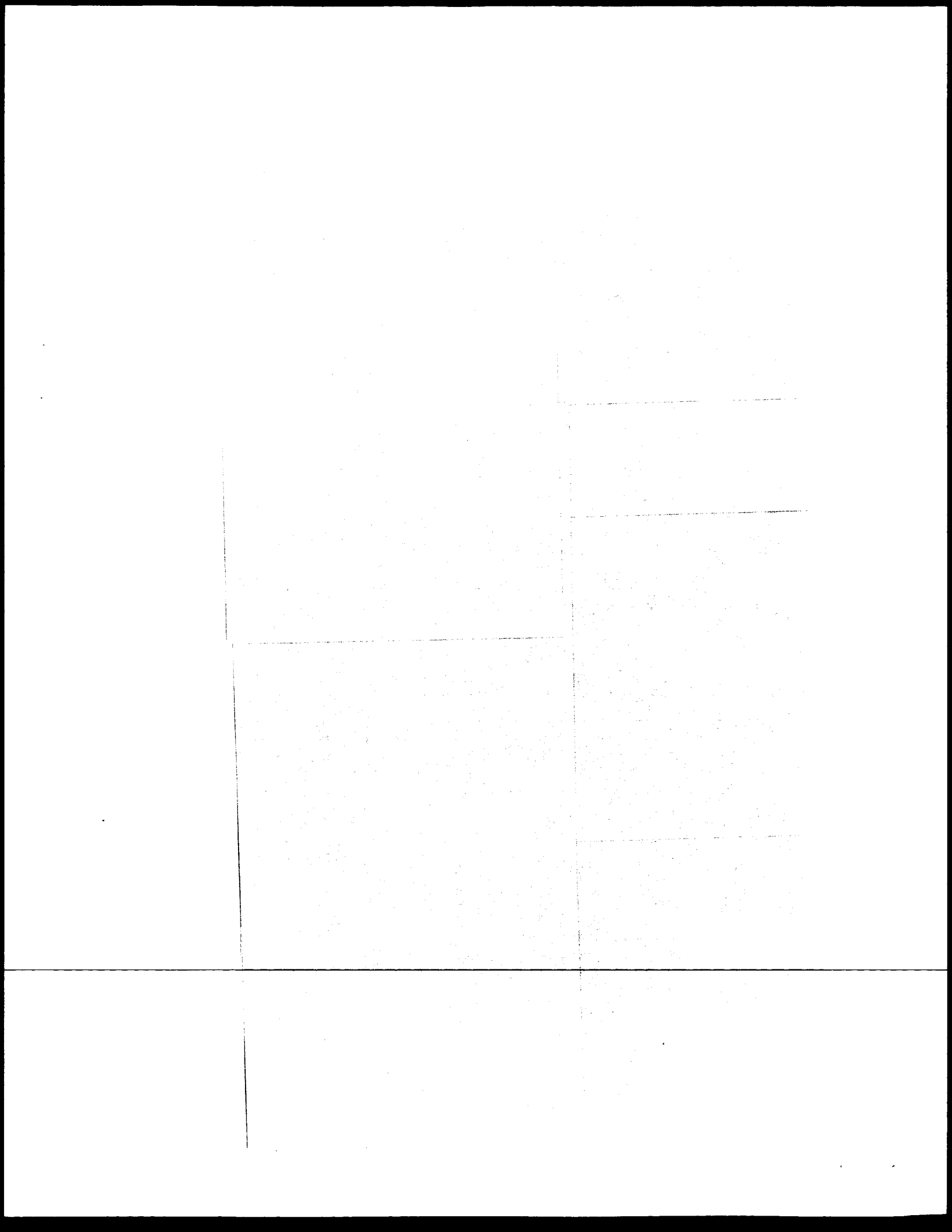
9 CONCLUSIONS

Molecular definition and preliminary characterization of a novel human heparanase paralog (heparanase II) has provided a new drug discovery target.

Figure 1







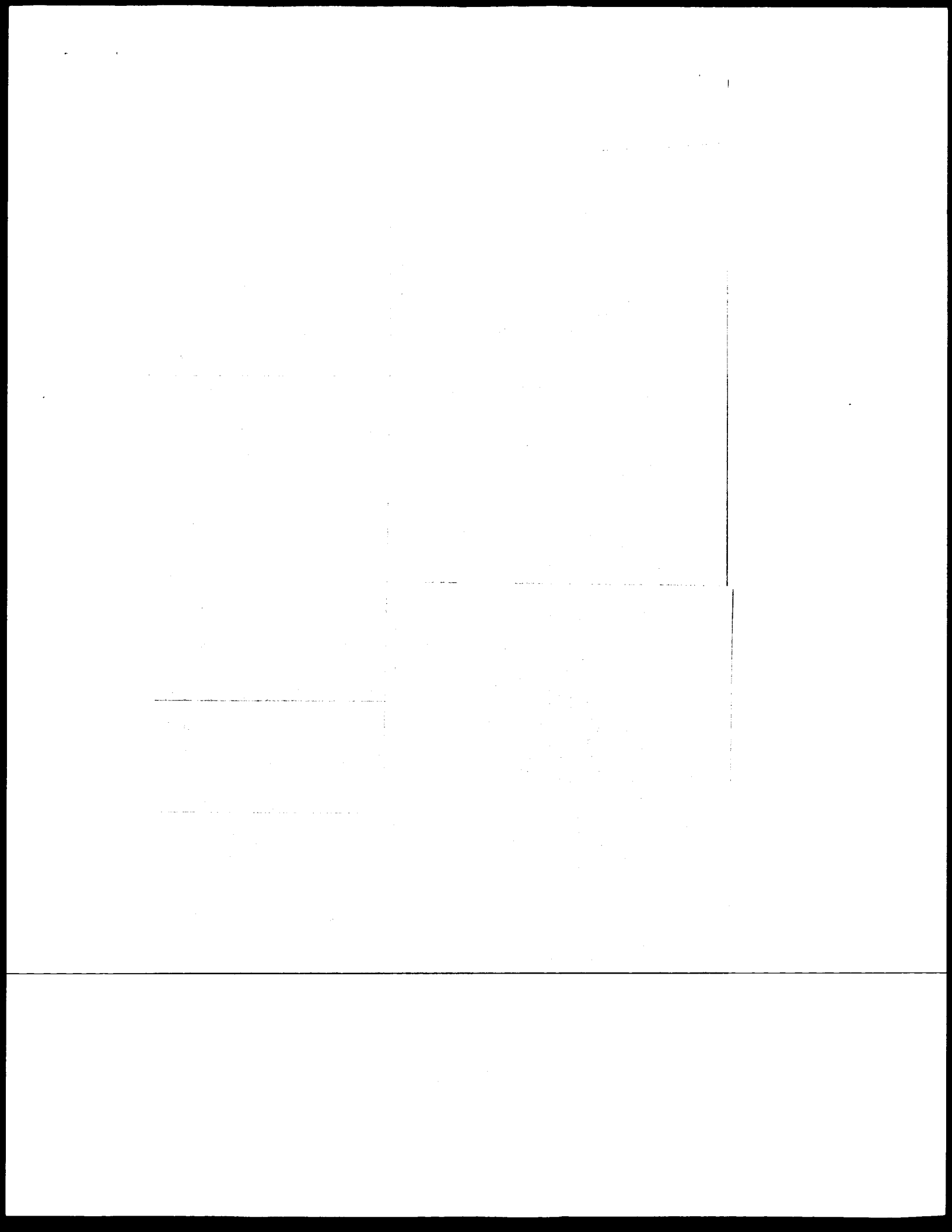


Figure 3 Composite cDNA and predicted amino acid sequence of human heparanase II

```
CAGGTTTTAAATCAGAGGGATTGAATGAGGGTGCTTTGTGCCTTCCCTGAAGCCATGCCC
                          M  R  V  L  C  A  F  P  E  A  M  P
TCCAGCAACTCCCGCCCCCCCGCGTGCCTAGCCCCGGGGGCTCTCTACTTGGCTCTGTTG
S  S  N  S  R  P  P  A  C  L  A  P  G  A  L  Y  L  A  L  L
CTCCATCTCTCCCTTTCCTCCCAGGCTGGAGACAGGAGACCCTTGCCTGTAGACAGAGCT
L  H  L  S  L  S  S  Q  A  G  D  R  R  P  L  P  V  D  R  A
GCAGGTTTGAAGGAAAAGACCCTGATTCTACTTGATGTGAGCACCAAGAACCCAGTCAGG
A  G  L  K  E  K  T  L  I  L  L  D  V  S  T  K  N  P  V  R
ACAGTCAATGAGAACTTCCTCTCTCTGCAGCTGGATCCGTCCATCATTCATGATGGCTGG
T  V  N  E  N  F  L  S  L  Q  L  D  P  S  I  I  H  D  G  W
CTCGATTTCCTAAGCTCCAAGCGCTTGGTGACCCTGGCCCGGGGACTTTCGCCCGCCTTT
L  D  F  L  S  S  K  R  L  V  T  L  A  R  G  L  S  P  A  F
CTGCGCTTCGGGGGCAAAAGGACCGACTTCCTGCAGTTCCAGAACCTGAGGAACCCGGCG
L  R  F  G  G  K  R  T  D  F  L  Q  F  Q  N  L  R  N  P  A
AAAAGCCGCGGGGGCCCGGGCCCGGATTACTATCTCAAAAACTATGAGGATGACATTGTT
K  S  R  G  G  P  G  P  D  Y  Y  L  K  N  Y  E  D  D  I  V
CGAAGTGATGTTGCCTTAGATAAACAGAAAGGCTGCAAGATTGCCCAGCACCCTGATGTT
R  S  D  V  A  L  D  K  Q  K  G  C  K  I  A  Q  H  P  D  V
ATGCTGGAGCTCCAAAGGGAGAAGGCAGCTCAGATGCATCTGGTTCTTCTAAAGGAGCAA
M  L  E  L  Q  R  E  K  A  A  Q  M  H  L  V  L  L  K  E  Q
TTCTCCAATACTTACAGTAATCTCATATTAACAGAGCCAAATAACTATCGGACCATGCAT
F  S  N  T  Y  S  N  L  I  L  T  E  P  N  N  Y  R  T  M  H
GGCCGGGCAGTAAATGGCAGCCAGTTGGGAAAGGATTACATCCAGCTGAAGAGCCTGTTG
G  R  A  V  N  G  S  Q  L  G  K  D  Y  I  Q  L  K  S  L  L
CAGCCCATCCGGATTTATTCCAGAGCCAGCTTATATGGCCCTAATATTGGGCGGCCGAGG
Q  P  I  R  I  Y  S  R  A  S  L  Y  G  P  N  I  G  R  P  R
AAGAATGTCATCGCCCTCCTAGATGGATTCATGAAGGTGGCAGGAAGTACAGTAGATGCA
K  N  V  I  A  L  L  D  G  F  M  K  V  A  G  S  T  V  D  A
GTTACCTGGCAACATTGCTACATTGATGGCCGGGTGGTCAAGGTGATGGACTTCCTGAAA
V  T  W  Q  H  C  Y  I  D  G  R  V  V  K  V  M  D  F  L  K
ACTCGCCTGTTAGACACACTCTCTGACCAGATTAGGAAAATTCAGAAAGTGGTTAATACA
T  R  L  L  D  T  L  S  D  Q  I  R  K  I  Q  K  V  V  N  T
TACACTCCAGGAAAGAAGATTTGGCTTGAAGGTGTGGTGACCACCTCAGCTGGAGGCACA
Y  T  P  G  K  K  I  W  L  E  G  V  V  T  T  S  A  G  G  T
AACAATCTATCCGATTCCTATGCTGCAGGATTCTTATGGTTGAACACTTTAGGAATGCTG
N  N  L  S  D  S  Y  A  A  G  F  L  W  L  N  T  L  G  M  L
GCCAATCAGGGCATTGATGTCGTGATACGGCACTCATTTTTTGACCATGGATACAATCAC
A  N  Q  G  I  D  V  V  I  R  H  S  F  F  D  H  G  Y  N  H
CTCGTGGACCAGAATTTTAACCCATTACCAGACTACTGGCTCTCTCTCCTCTACAAGCGC
L  V  D  Q  N  F  N  P  L  P  D  Y  W  L  S  L  L  Y  K  R
CTGATCGGCCCCAAAGTCTTGGCTGTGCATGTGGCTGGGCTCCAGCGGAAGCCACGGCCT
L  I  G  P  K  V  L  A  V  H  V  A  G  L  Q  R  K  P  R  P
GGCCGAGTGATCCGGGACAAACTAAGGATTTATGCTCACTGCACAAACCACCACAACCAC
G  R  V  I  R  D  K  L  R  I  Y  A  H  C  T  N  H  H  N  H
AACTACGTTCGTGGGTCCATTACACTTTTTATCATCAACTTGCATCGATCAAGAAAGAAA
N  Y  V  R  G  S  I  T  L  F  I  I  N  L  H  R  S  R  K  K
ATCAAGCTGGCTGGGACTCTCAGAGACAAGCTGGTTCACCAGTACCTGCTGCAGCCCTAT
I  K  L  A  G  T  L  R  D  K  L  V  H  Q  Y  L  L  Q  P  Y
GGGCAGGAGGGCCTAAAGTCCAAGTCAGTGCAACTGAATGGCCAGCCCTTAGTGATGGTG
G  Q  E  G  L  K  S  K  S  V  Q  L  N  G  Q  P  L  V  M  V
GACGACGGGACCCTCCCAGAATTGAAGCCCCGCCCCCTTCGGGCCGGCCGGACATTGGTC
D  D  G  T  L  P  E  L  K  P  R  P  L  R  A  G  R  T  L  V
ATCCCTCCAGTCACCATGGGCTTTTTTGTGGTCAAGAATGTCAATGCTTTGGCCTGCCGC
I  P  P  V  T  M  G  F  F  V  V  K  N  V  N  A  L  A  C  R
TACCGATAAGCTATCCTCACACTCATGGCTACCAGTGGGCCTGCTGGGCTGCTTCCACTC
Y  R
CTCCACTCCAGTAGTATCCTCTGTTTTCAGACATCCTAGCAACCAGCCCCTGCTGCCCCA
TCCTGCTGGAATCAACACAGACTTGCTCTCCAAAGAGACTAAATGTCATAGCGTGATCTT
AGCCTAGGTAGGCCACATCCATCCCAAAGGAAAATGTAGACATCACCTGTACCTATATAA
GGATAAAGGCATGTGTATAGAGCAGAATGTTTCTCTTCATGTGCACTATGAAAACGAGCT
GACAGCACACTCCCAGGAGAAATGTTTCCAGACAACTCCCCATGATCCTGTCACACAGCA
TTATAACCACAAATCCAAACCTTAGCCTGCTGCTGCTGCTGCCCTCAGAGGAAGATGAGG
AAGGAAAAAAAACTGGGTGGACCTACAAAAACCCATCCTCTCCCAACTCCTTCTTCTCTG
CCTCTTTCTTGCTGCTGCCCTGAGTTTTTTGACACATCTCTTTCCATAGGGGAGTAATGG
GTGTGTCAGCCCTGGCCTGCTGGGAGAGCTGTTTATATGATTTCCCGGCTGATGTATGAG
CGTGCGCACCTGGGTTCCTCACAGTGGCATCCATCACTGGCAGTTCTTCTGGGAAGCGGG
TGCTTCAAAAGTAAAATTACAATCACACTCCAAAAAAAAAAAAAAA
```

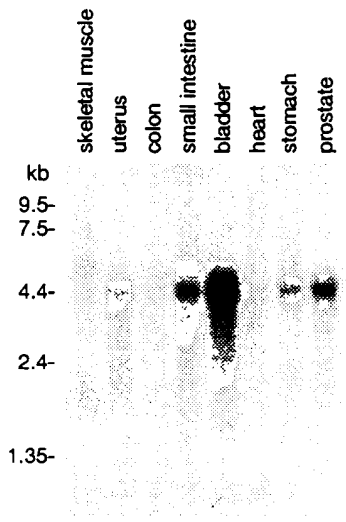
Figure 4 Predicted amino acid sequence of human heparanase II depicting functional motifs. The signal peptide is shown in bold, canonical acceptor sites for N-linked glycosylation are double underlined, and predicted sites for phosphorylation by protein kinase C are underlined.

MRVLCAFPEAMPSSNSRPPACLAPGALYLALLLHLSLSSQAGDRRPLPVDRAAGLKEKTL
ILLDVSTKNPVRTVNENFLSLQLDPSIIHDGWLDFLSSKRLVTLARGLSPAFLRFGGKRT
DFLQFQNLRNPAKSRGGPGPDYYLKNYEDDIVRSDVALDKQKGCKIAQHPDVMLELQREK
AAQMHLVLLKEQFSNTYSNLILTEPNNYRTMHGRAVNGSQLGKDYIQLKSLLQPIRIYSR
ASLYGPNIGRPRKNVIALLDGFMKVAGSTVDAVTWQHCYIDGRVVKVMDFLKTRLLDTLS
DQIRKIQKVVNTYTPGKKIWLEGVVTTSAGGTNNLSDSYAAGFLWLNTLGMLANQGIDVV
IRHSFFDHGYNHLVDQNFNPLPDYWLSLLYKRLIGPKVLAVHVAGLQRKPRPGRVIRDKL
RIYAHCTNHHNHNYVRGSITLFIINLHRSRKKIKLAGTLRDKLVHQYLLQPYGQEGLKSK
SVQLNGQPLVMVDDGTLPELKPRPLRAGRTLVIPPVTMGFFVVKNVNALACRYR

Figure 5 Clustal W alignment of human heparanase I and human heparanase II

```
HEPI    --MLLRSKPALP-----PP-------LMLLLLGPLGPLSPGALPRPA--------QAQDV
HEPII   MRVLCAFPEAMPSSNSRPPACLAPGALYLALLLHLSLSSQAGDRRPLPVDRAAGLKEKTL
          :*     *:*     **        * * **  *.  * ..  **          : : :

HEPI    VDLDFFTQEPLHLVSPSFLSVTIDANLATDPRFLILLGSPKLRTLARGLSPAYLRFGGTK
HEPII   ILLDVSTKNPVRTVNENFLSLQLDPSIIHDG-WLDFLSSKRLVTLARGLSPAFLRFGGKR
        : **. *::*:: *. .***: :*..:  *  :* :*.* :* *********:*****.:

HEPI    TDFLIFDPKKESTFEERSYWQSQVNQDICKYGSIPPDVEEKLRLEWPYQEQLLLREHYQK
HEPII   TDFLQFQNLRN---PAKSR------------GGPGPDYYLKNYEDDIVRSDVALDK--QK
        **** *:  ::     :*             *.  **   *   :   :.:: * :  **

HEPI    KFKNSTYSRSSVDVLYTFANCSGLDLIFGLNALLRTADLQWNSSNAQLLLDYCSSKGYNI
HEPII   GCKIAQHP----DVMLELQREK------AAQMHLVLLKEQFSNTYSNLIL---T------
          * : :.    **:  : . .      . :  *   . *:..: ::*:*   :

HEPI    SWELGNEPNSFLKKADIFINGSQLGEDFIQLHKLLRKS-TFKNAKLYGPDVGQPRRKTAK
HEPII   ---E---PNNYRTMHGRAVNGSQLGKDYIQLKSLLQPIRIYSRASLYGPNIGRPRKNVIA
              **.: .  .  :******:*:***:.**:    :..*.****::*:**::.

HEPI    MLKSFLKAGGEVIDSVTWHHYYLNGRTATKEDFLNPDVLDIFISSVQKVFQVVESTRPGK
HEPII   LLDGFMKVAGSTVDAVTWQHCYIDGRVVKVMDFLKTRLLDTLSDQIRKIQKVVNTYTPGK
        :*..*:*..*..:*:***:* *::**...  ***:. :** : ..::*: :**::  ***

HEPI    KVWLGETSSAYGGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDEN
HEPII   KIWLEGVVTTSAGGTNNLSDSYAAGFLWLNTLGMLANQGIDVVIRHSFFDHGYNHLVDQN
        *:**  . :: .**:  ***::****:**:.**: *. **:**:*: **. *  ****:*

HEPI    FDPLPDYWLSLLFKKLVGTKVLMASVQGSKRR---------KLRVYLHCTNTDNPRYKEG
HEPII   FNPLPDYWLSLLYKRLIGPKVLAVHVAGLQRKPRPGRVIRDKLRIYAHCTNHHNHNYVRG
        *:**********:*:*:*.*** . * * :*:         ***:* **** .* .* .*

HEPI    DLTLYAINLHNVTKYLRLPYPFSNKQVDKYLLRPLGPHGLLSKSVQLNGLTLKMVDDQTL
HEPII   SITLFIINLHRSRKKIKLAGTLRDKLVHQYLLQPYGQEGLKSKSVQLNGQPLVMVDDGTL
        .:**: ****.  * ::*. .: :* *.:***:* * .** ******** .* **** **

HEPI    PPLMEKPLRPGSSLGLPAFSYSFFVIRNAKVAACI--
HEPII   PELKPRPLRAGRTLVIPPVTMGFFVVKNVNALACRYR
        * *  :***.* :* :*..: .***::*.:. **
```

Heparanase 2 Northern Blot



10 REFERENCES

1. Bar-Ner, M., Mayer, M., Schirrmacher, V., and Vlodavsky, I. "Involvement of both heparanase and plasminogen activator in lymphoma cell-mediated degradation of heparan sulfate in the subendothelial extracellular matrix" *J. Cell. Biol.* 128:299-306 (1986)
2. Yurchenco, P.D. and Schittny, J.C. "Molecular architecture of basement membranes" *FASEB J.* 4:1577-1590 (1990)
3. Taipale, J. and Keski-Oji, J. "Growth factors in the extracellular matrix" *FASEB J.* 11:51-59 (1997)
4. Vlodavsy, I, Eldor, A., Haimovitz, F.A., Matzner, Y., Ishai, M.R., Lider, O., Naparstek, Y., Cohen, I.R., and Fuks, Z. "?????"*Invasion Metastatsis* 12:112-127 (1992)
5. Vlodavsky, I., Mohsen, M. Lider, O., Ishai-Michaeli, R., Ekre, H.P., Svanhn, C.M., Vogoda, M., and Peretz, T. "???"*Invasion Metastatsis* 14:290-302 (1995)
6. Bashkin, P., Doctrow, S., Klagsbrun, M., Svahn, C.M., Folkman, J., and Vlodavsky, I. "????" *Biochemistry* 28:1737-1743 (1989)
7. Vlodavsky, I., Friedmann, Y., Elkin, M., Aingorn, H., Atzmon, R., Ishai-Michaeli, R., Bitan, M., Pappo, O., Peretz, T., Michal, I., Spector, L., and Pecker, I. "Mammalian heparanase: Gene cloning, expression and function in tumor progression and metastasis" *Nature Genetics* 5:793-802 (1999)
8. Hulett, M.D., Freeman, C., Hamdorf, B.J., Baker, R., Harris, M.J., and Parish, C.R. "Cloning of mammalian heparanase, an important enzyme in tumor invasion and metastasis" *Nature Genetics* 5:803-809 (1999)
9. Toyoshima, M. and Nakajima, M. "Human heparanase: Purification, characterization, cloning and expression" *J. Biol. Chem.* 274:24153-24160 (1999)
10. Kussie, P.H., Hulmes, J.D., Ludwig, D.L., Patel, S., Navarro, E.C., Seddon, A.P., Giorgio, N.A., and Bohlen, P. "Cloning and functional expression of a human heparanase gene" *Biochem. Biophys. Res. Comm.* 261:183-187 (1999)
11. Fairbanks, M.B., Mildner, A.M., Leone, J.W., Cavey, G.S., Mathews, W.R., Drong, R.F., Slightom, J.L., Bienkowski, M.J., Smith, C.W., Bannow, C.A., and Heinrikson, R.L. "Processing of the human heparanase precursor and evidence that the active enzyme is a heterodimer" *J. Biol. Chem.* 274:29587-29590 (1999)
12. Nielsen, H. Engelbrecht, J. , Brunak, S., and von Heijne, G. "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites" *Protein Engineering* 10:1-6 (1997)
13. Sonnhammer, E.L., von Heijne, G., and Krogh, A. "A hidden Markov model for predicting transmembrane helices in protein sequences" In *Proc. of Sixth Int. Conf. on Intelligent Systems for Molecular Biology*, p 175-182 Ed J. Glasgow, T. Littlejohn, F. Major, R. Lathrop, D. Sankoff, and C. Sensen Menlo Park, CA: AAAI Press (1998)
14. Hansen, J.E., Lund, O., Tolstrup, N., Gooley, A.A., Williams, K.L., and Brunak, S. "NetOGlyc: Prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility" *Glycoconjugate Journal* 15: 115-130 (1998)

15. Pikas, D.S., Li, J., Vlodavsky, I., and Lindahl, U. "Substrate specificities of heparanases from human hepatoma and platelets" *J. Biol. Chem.* 273:18770-18777 (1998)
16. McCarter, J.D. and Withers, S. G., "" *Curr. Opin. Struct. Biol.* 4:885-892 (1994)
17. Sinnott, M.L. "Catalytic mechanisms of enzymic glycosyl transfer" *Chem. Rev.* 90:1171-1202 (1990)
18. Wong, A.W., He, S., Grubb, J.H., Sly, W.S., and Withers, S.G. "Identification of Glu-540 as the catalytic nucleophile of human β-glucuronidase using electrospray mass spectrometry" *J. Biol. Chem.* 273:34057-34062 (1998)
19. Islam, M.R., Tomatsu, S., Shah, G.N., Grubb, J.H., Jain, S., and Sly, W.S. "Active site residues of human β-glucuronidase" *J. Biol. Chem.* 274:23451-23455 (1999)
20. Phillips. D.C. "The three dimensional structure of an enzyme molecule" *Sci. Amer.* 215:79-90 (1966)
21. Freeman, C., Browne, A.M., and Parish, C.R. "Evidence that platelet and tumor heparanases are similar enzymes" *Biochem. J.* 342:361-368 (1999)
22. Freeman, C. and Parish, C.R. "A rapid quantitative assay for the detection of mammalian heparanase activity" *Biochem J.* 325:229-237 (1997)